CHAPTER 5

Relationships between variables

In this chapter we will look at different sorts of relationships. By considering the different sorts of measurement scales (see paragraph 2.1), we obtain the following relationships between combinations of the measurement scales:

- Linear relationships between two continuous (interval / ratio scaled) variables;
- Linear relationships between one continuous variable and more than one independent continuous variable, which can be ordinal or dichotomous;
- Relationship between a continuous and a dichotomous variable;
- Relationship between two dichotomous variables;
- Relationship between two nominal variables.

5.1 Effect size of linear relationships between two continuous variables

The Pearson Product moment correlation coefficient, ρ_{xy}, between the continuous variables x and y obtained from population elements, is a measure of the linear relationship between x and y. The index ρ_{xy} is dimensionless and takes on values between -1 and 1, where the values 1 and -1 indicate a perfect linear relationship and perfect inverse linear relationship between x and y. When $\rho_{xy} = 0$ it means that there is no linear relationship. The effect size-index does not depend on measurement units and so Cohen (1969, 1977, 1988) recommends ρ_{xy} (abbreviated by ρ) as the effect size-index. Since $z_y = \rho\, z_x$, with z_x and z_y the standard scores of x and y, ρ can be considered as the number of standard deviation units that y increases as x increases one standard deviation unit.

Example 5.1 (Rothmann et.al, 2000a):
As in Example B, Chapter 3, the MBTI is also applied to pre-graduate pharmacy students at a university in order to determine the relationship between academic achievement and personality preference scores. Table 5.1 displays the results of the correlation between academic achievement and E/I preference scores. It is clear from this table that the effect sizes of linear relationships (ρ), especially in males, decreases in their 2nd and 3rd years, but increases sharply after that. The relationships are mostly far from perfect $(\rho = 1)$ and are actually weak relationships $(\rho$ close to $0)$.

Table 5.1

Correlations between academic achievement and E/I – preference scores

	Academic year			
	1	2	3	4
ρ (males)	0,23	0,13	-0,05	0,47
ρ (females)	0,24	0,15	0,20	0,34

When random samples are drawn from a population, ρ can be estimated by the sample correlation coefficient r. This estimation is unfortunately biased for ρ, with the bias approximately equal to $-\frac{1}{2}\rho\left(1-\rho^2\right)/n$, which always lies between $-0{,}2/n$ and $0{,}2/n$ (see Steyn, 2002). This means that, for large samples, r is an unbiased estimator, but for small n, r underestimates positive ρ values and overestimates negative ρ values. Grissom & Kim (2005: 72) give the approximate unbiased estimator for ρ as:

$$\hat{r} = r + \frac{r(1-r^2)}{2(n-3)}. \qquad (5.1)$$

Example 5.2 (Steyn, 2002)**:**

The inter-correlations between 6 aptitude scores of a random sample of 112 people is summarized in Table 5.2.

With the exception of the correlation of 0,184 between maze navigation ability and reading comprehension, all of the correlations were statistically significant at a 5% level. This only means that the correlations can be considered to be non-zero. The estimated effect sizes, r, are, on average, no more than 0,2/112 = 0,0018 smaller than the population-effect sizes ρ, and thus reasonably accurate estimations. According to (5.1) the approximate unbiased estimation of the correlation between Non-verbal intelligence and Picture completion is:: $\hat{r} = 0,466 + \frac{0,466(1-0,466^2)}{2(112-3)} = 0,466 + 0,0017 = 0,467.$

Table 5.2

Correlation coefficient of aptitude test scores

Skill	**1**	**2**	**3**	**5**	**6**
1. Non-verbal intelligence					
2. Picture completion	0,466**				
3. Block design	0,552**	0,572**			
4. Maze navigation	0,340*	0,193	0,445**		
5. Reading comprehension	0,576**	0,263*	0,354*	0,184	
6. Vocabulary	0,510**	0,239*	0,356*	0,219*	0,794**

* Medium effect. ** Large effect

5.1.1 Guideline values for correlation effect size indices

If the correlations are used as effect sizes, the question must once again be asked: "How large must these values be before they indicate an important relationship?" Cohen (1969, 1977, 1988) proposes the following guideline values:

- Small effect: $|\rho| = 0{,}1$
- Medium effect: $|\rho| = 0{,}3$
- Large effect: $|\rho| = 0{,}5$.

He motivates these values briefly as follows:

(a) Small effect: $|\rho| = 0{,}1$ is a small correlation and means that only 1% (i.e., $100 \times \rho^2 = 100 \times 0{,}1^2$) of x's variance is explained by y.

(b) Medium effect: $|\rho| = 0{,}3$ is a correlation which is typically found in behavioural sciences (see e.g., Example 5.1). These relationships are observable by inspection (they can be "eye-balled"). In psychometric tests' determination of validity using a criterion, one expects to find correlations between 0 and 0,6, with the majority of them lower than 0,3 (this follows from a statement made by Guilford (1965: 146).

(c) Large effect: $|\rho| = 0{,}5$ means that x explains 25% of y's variance, so that x and y are clearly linearly related. According to a statement made by Ghiselli (1964: 61), 0,5 is a practical upper-bound for correlations obtained for validity calculations, which does not drastically differ from Guilford's 0,6 discussed above. While correlations between IQ or other similar tests concerning scholastic achievement vary about 0,5, correlations between personality measures and comparable criteria are more likely to vary about 0,3 (which usually indicates a medium effect).

Feinstein (1999: 2569) provides different guidelines than those provided by Cohen. His reasoning is as follows: In a typical regression context with response variable *y* and predictor variable *x*, many data-analysts (such as Fleiss, 1981: 60 and Burnand et al., 1990) do not agree that *x* effectively explains the variation in *y* unless the percentage of variation accounted for by *x* exceeds 10%. This means that we need $\mathrm{r}^2 \geq 0{,}1$ or approximately that $\mathrm{r} \geq 0{,}3$.

In paragraph 5.3.1 we want to reconcile these effect size indices with the standardized difference δ in order to provide a further motivation for the guideline values.

In Table 5.2 the medium and large effects are highlighted. Note that the guidelines are not rigidly applied, but are rather used to indicate a "region" for classification. For this reason 0,466 is taken as a large effect and 0,239 and 0,354 are taken as medium effects, since they are respectively close to 0,5 and 0,3.

5.1.2 Confidence intervals for correlation effect sizes

Under the assumption that the x and y form a bivariate normal distribution with correlation coefficient ρ, then the statistic

$$z(r) = \frac{1}{2}\ell n\left(\frac{1+r}{1-r}\right) \tag{5.2}$$

has an *approximate normal* distributed with mean $z(\rho)$ and variance $1/(n-3)$, where n is the size of a random sample and r is the sample correlation coefficient (see Snedecor & Cochran, 1980: 186).

The $100(1-\alpha)\%$ *CI* for $z(\rho)$ has the following boundaries

$$z(\rho_L) = z(r) - z_{\alpha/2}/\sqrt{n-3}$$

$$z\left(\rho_U\right) = z\left(r\right) + z_{\alpha/2} / \sqrt{n-3} \tag{5.3}$$

The following result can then be derived from (5.2)

$$r = \frac{e^{2z(r)} - 1}{e^{2z(r)} + 1} \quad , \tag{5.4}$$

and then lower and upper bounds for the $100\left(1 - \alpha\right)\%\ CI$ for ρ are:

$$\rho_L = \left(e^{2z(\rho_L)} - 1\right) / \left(e^{2z(\rho_L)} + 1\right)$$

and (5.5)

$$\rho_U = \left(e^{2z\left(\rho_U\right)} - 1\right) / \left(e^{2z\left(\rho_U\right)} + 1\right) .$$

Example 5.3:

In Example 5.2 the 95% *CI* of the correlation between Non-verbal intelligence and Picture completion is calculated as follows:

$$z\left(0{,}466\right) = \tfrac{1}{2}\ell n\left(\frac{1+0{,}466}{1-0{,}466}\right) = \tfrac{1}{2}\ell n\left(2{,}745\right) = 0{,}505$$

$$z\left(\rho_L\right) = 0{,}505 - 1{,}96 / \sqrt{112-3} = 0{,}505 - 0{,}188 = 0{,}317$$

$$z\left(\rho_U\right) = 0{,}505 + 0{,}188 = 0{,}693$$

$$\rho_L = \left(e^{2\times 0{,}317} - 1\right) / \left(e^{2\times 0{,}317} + 1\right)$$

$$= \left(1{,}885 - 1\right) / \left(1{,}885 + 1\right) = 0{,}307.$$

$$\rho_U = \left(e^{2\times 0{,}693} - 1\right) / \left(e^{2\times 0{,}693} + 1\right) = \left(3{,}999 - 1\right) / \left(3{,}999 + 1\right) = 0{,}600$$

This suggests that the population effect size index for the linear relationship between the aptitude scores mentioned can be as small as 0,307 (medium effect), but as large as 0,600 (large effect), with a 95% probability.

Note that the *CI* is not symmetric about the estimated value 0,466– which is different from the *CI*'s for standardized differences.

5.1.3 Counternull values for correlation

Because z(r) has an approximate normal distribution with mean z(ρ) and variance 1/(n-3), the counternull value of z(r) is:

$$r_{conternull} = \frac{e^{2z_{counternull}} - 1}{e^{2z_{counternull}} + 1}.$$

Therefore, because the distribution of r is not symmetric and the variance is a function of r, we make use of Fisher's z-transformation which has a symmetric distribution with variance independent of r. The back-transformation of 2z(r) produces the required counter null value for r.

Example 5.3 (continued)

With r = 0,466 and z(r) = 0,505 the null-counternull value of $z_{counternull} = 2\times 0,505 = 1,1$ is

$$\mathrm{r}_{\mathrm{counternull}} = \frac{\mathrm{e}^{2\times 1,1} - 1}{\mathrm{e}^{2\times 1,1} + 1} = \frac{9,03 - 1}{9,03 + 1} = 0,80.$$

This means that for r = 0,466 (p < 0,01) the correlation ρ could just as easily be as large as 0,8 as 0 and that the null-counternull interval (0,0 ; 0,8) can be interpreted as a 99% CI (i.e., $1-\alpha > 1-0,01 = 0,99$) for ρ .

5.1.4 Modification of correlation for reliability

Suppose that the measurement x can be written as $x = T + e_x$, where T is the true *score or measurement* and e_x is an error term. Similarly, y can be written as $y = U + e_y$. The *reliability* of x is defined as

$$\rho_{xx} = \frac{Var(T)}{Var(x)}$$, from which follows: $Var(T) = \rho_{xx} Var(x)$

and for y as

$\rho_{yy} = \frac{Var(U)}{Var(y)}$, from which follows: $Var(U) = \rho_{yy} Var(y)$.

This means that if the errors e_x and e_y are small relative to T and U, then ρ_{xx} and ρ_{yy} will be close to 1. Also, if the errors are very large, the reliability's value will be close to zero. For sample measurements, the values ρ_{xx} and ρ_{yy} can be estimated by r_{xx} and r_{yy} (see Steyn, 2004). If ρ_{xx} is the population correlation and r_{xy} is the sample correlation between x and y, then, according to Hunter & Schmidt, (2004: 96), the correlation between the true scores T and U can be given by

$$\rho_{TU} = \frac{Kov(T,U)}{\sqrt{Var(T)}\sqrt{Var(U)}} = \frac{Kov(x,y)}{\sqrt{\rho_{xx}Var(x)}\sqrt{\rho_{yy}Var(y)}} = \frac{\rho_{xy}}{\sqrt{\rho_{xx}\rho_{yy}}} \tag{5.6}$$

which can be estimated by

$$r_{TU} = \frac{r_{xy}}{\sqrt{r_{xx}}\sqrt{r_{yy}}} \quad . \tag{5.7}$$

This modification is commonly known as the *attenuation correction*.

Notes:

- Note that ρ_{TU} and r_{TU} are larger than ρ_{xy} and r_{xy} .
- If only y contains an error term and x is fixed, then

  $$\rho_{TU} = \frac{\rho_{xy}}{\sqrt{\rho_{yy}}} \quad \text{and} \quad r_{TU} = \frac{r_{xy}}{\sqrt{r_{yy}}} \tag{5.8}$$

  because in this case $r_{xx} = 1$, since $x = T$.
- The standard error of ρ_{TU} and r_{TU} is larger than that of ρ_{xy} and r_{xy} (Hunter & Schmidt, 2004: 96).
- If we want to test the reliability of a certain item in a psychometric test and the test's reliability is known, then the attenuation correction can be used as follows (Wanous en Hudy, 2001):

Let S denote the total count of items $X_1, X_2, ..., X_k$ with reliability ρ_{ss} and let the correlation between X_i and S be equal to ρ_{iS}. Because X_i and S measure the same underling property, we can assume that the correlation between the true X_i and S observations is equal to 1, such that, from (5.6):

$$1 = \frac{\rho_{iS}}{\sqrt{\rho_{ii}}\sqrt{\rho_{SS}}},$$

And the reliability of X_i is given by:

$$\rho_{ii} = \frac{\rho_{iS}^2}{\rho_{SS}}, \tag{5.9}$$

with estimator

$$r_{ii} = \frac{r_{iS}^2}{r_{SS}} \ . \tag{5.10}$$

In the case where the correlation between the two true observations of X_i and S is found to be less than 1, ρ_{ii} becomes larger. Values from (5.9) and (5.10) can thus be used as the **lower bound** for ρ_{ii} and r_{ii}.

Example 5.4:

In Example A, Chapter 3, the inter-correlations of the before test scores of the BDI and POMS_A and POMS_D are given as follows:

	POMS_A(x_2)	POMS_D(x_3)
BDI(x_1)	0,38	0,49
POMS_A(x_2)		0,73

From de Klerk et al. (2004) it is known that $r_{x_1x_1} = 0,84$, $r_{x_2x_2} = 0,82$ and $r_{x_3x_3} = 0,89$

are the “reliabilities” for BDI, POMS_A and POMS_D respectively. Thus, the correlations between the true scores are:

for BDI versus POMS_A : $r_{TU} = \frac{0,38}{\sqrt{0,84}\sqrt{0,82}} = 0,45$

BDI versus POMS_D : $r_{TU} = \frac{0,49}{\sqrt{0,84}\sqrt{0,89}} = 0,57$

POMS_A versus POMS_D : $r_{TU} = \frac{0,73}{\sqrt{0,82}\sqrt{0,89}} = 1,17$

Because a correlation cannot be larger than 1, we set $r_{TU} = 1$ in the last case.

5.2 Effect sizes of linear relationships between a continuous response variable and more than one predictor variable

In multiple linear regression the linear relationship between a response variable y and predictors $x_1, x_2, ..., x_u$ is determined by the multiple correlation coefficient $R_{y.A}$, where A represents the set of predictor variables. The square of the multiple correlation coefficient of $R_{y.A}$, namely $R_{y.A}^2$, the multiple coefficient of determination is obtained, which describes the *proportion variance of* $\boldsymbol{y}$ *which is explained by the multiple regression-relationship*

$$\hat{y} = a + b_1x_1 + b_2x_2 + ... + b_ux_u \quad \text{, i.e.,}$$

$$R_{y.A}^2 = \sum(\bar{y} - \hat{y})^2 / \sum(y - \bar{y})^2 = 1 - \left[\sum(y - \hat{y})^2 / \sum(y - \bar{y})^2\right],$$

where $\bar{y}$ is the mean of the y variable. This proportion variance can now serve as an effect size-index.

Usually $R^2_{y.A}$ is defined in terms of the random sample values y and $x_1, \ldots, x_k$, and so, for the population-case, it will be denoted by $\rho^2_{y.A}$.

5.2.1 Semi-partial R^2 as an effect size-index

Another statistic which can be used as an effect size index and is used extensively in multiple linear regression, is the proportion of y's variance which is explained by a set of predictors B, apart from that explained by another set A. Cohen (1969, 1977, 1988) defines it as

$$R^2_{y.A,B} - R^2_{y.A}, \tag{5.11}$$

and it is commonly known as the *squared semi-partial multiple correlation* (see also Smithson, 2001). This quantity can be interpreted as the *contribution* that the predictors in B make to R^2 in a regression which includes all of the predictors of both A and B. It is called a semi-partial R^2 because A's influence on B is removed, but A's influence on y is still considered. The population analogue for semi- partial R^2 is denoted by $\rho^2_{y.A,B} - \rho^2_{y.A}$.

5.2.2 Partial R^2 as an effect size-index

If A's influence on B and y is removed, then we obtain

$$R^2_{yB.A} = \frac{R^2_{y.A,B} - R^2_{y.A}}{1 - R^2_{y.A}}, \tag{5.12}$$

the *partial* R^2 (Cohen, 1969, 1977, 1988).

In this case the semi-partial R^2 expressed as a proportion of the portion of the variance of y not explained by A (i.e., $1 - R^2_{y.A}$) is not provided.

5.2.3 The effect size index f^2

Cohen (1969, 1977, 1988) defines the index $f^2 = PV_B / PV_F$, which is simply the ratio of the $PV's$, i.e., it is the ratio of the proportion variance of y explained by some source B, to the proportion variance, PV_F, of error or residual variance. It can also be seen as the signal-noise ratio in a multiple regression context. When we only make use of the set of predictors A, the statistic is as follows

$$f^2 = R^2_{y.A} / \left(1 - R^2_{y.A}\right) \tag{5.13}$$

The proportion variances explained, namely $R^2_{y.A}$, $R^2_{y.A,B} - R^2_{y.A}$ and $R^2_{yB.A}$, are all easily interpreted because they all lie between 0 and 1, i.e., as these values become larger, it becomes clear that the variables contained within the multiple regression model are more capable of explaining the variance, or information, contained within y than those variables which were excluded from the model. For these reasons, the proportion variances mentioned serve as better effect size indices than f^2, and so we will not concern ourselves with f^2 any further.

5.2.4 Guideline values for proportion variance

The statistic $R^2_{y.A}$ is a generalization of $r^2_{x,y}$ and the parameter $\rho^2_{y.A}$ is a generalization of ρ^2_{xy}, consequently sensible guideline values are:

- Small effect: $\rho^2_{y.A} = 0{,}01,$ or $\left|\rho_{y.A}\right| = 0{,}1,$ which was a small effect for ρ_{xy}. This means that only 1% of y's variance is explained by the regression based on the predictors in the set *A*.

- Medium effect: $\rho^2_{y.A} = 0,1$, or $|\rho_{y.A}| = 0,317$, which is roughly a medium effect when considering ρ_{xy}. In this case 10% of y's variance is explained. Feinstein (1999: 2569) provides different cut-off points for a "significant" effect.
- Large effect: $\rho^2_{y.A} = 0,25$, or $|\rho_{y.A}| = 0,5$, which would have been considered large for ρ_{xy}.

Example 5.5 (Smithson, 2001: 616):

Suppose that the number of visits to professional health services (y) is predicted from measures of psychological health (x_1), physical health (x_2) and stress level (x_3) by making use of a multiple linear regression. Let $A = \{x_1\}$ and $B = \{x_2, x_3\}$. A sample of 465 people is used to fit the multiple linear regression model, and it produces the following results:

$$r^2_{yx1} = R^2_{y.A} = 0{,}1261, \text{ while } R^2_{y.A,B} = 0{,}3768.$$

The proportion variance of y explained by x_1 is 0,1261, which indicates a medium effect, while x_1, x_2 and x_3 together explain 0,3768 of the variance. This can be considered a large effect.

The semi partial R^2 is

$$R^2_{y.A,B} - R^2_{y.A} = 0{,}3768 - 0{,}1261 = 0{,}2507,$$

so that the proportion variance of y explained by x_2 and x_3 without x_1 being considered, has a large effect.

Finally the partial R^2 of y with x_2 and x_3, when x_1's influence is removed, is:

$$R^2_{yB.A} = 0{,}2507/(1 - 0{,}1261) = 0{,}2869,$$

which also indicates a large effect.

5.2.5 Point and interval estimation of proportion variance (Smithson, 2001)

If $R^2_{y.A}$ is used to estimate $\rho^2_{y.A}$, then it is positively biased for small samples. An unbiased estimator is the *modified* R^2, defined as:

$$R_a^2 = R^2 - \left(1 - R^2\right)\left(\frac{u}{v}\right) \qquad (5.14)$$

where $R^2 = R^2_{y.A}$, *u* is the number of predictors in A, and $v = n - u - 1$. The test statistic to test the null hypothesis $H_0 : \rho^2_{y.A} = 0$, is the F- statistic defined as

$$F(u,v) = \frac{R^2_{y.A} / u}{\left(1 - R^2_{y.A}\right) / v} \quad , \qquad (5.15)$$

and if H_0 is not necessarily true, and the population distribution of y (given the predictors in A) is normal, then $F(u,v)$ follows a non-central F-distribution with non-centrality parameter

$$ncp = \left(\frac{\rho^2_{y.A}}{1 - \rho^2_{y.A}}\right)\left(u + v + 1\right) \quad . \qquad (5.16)$$

Similarly to the methods discussed in paragraph 4.1.2, an exact $100(1-\alpha)$% *CI* for ncp can be determined by making use of computer software packages where the only input required for these programs are the values u, v, $F(u,v)$ (defined in (5.15)) and α. From this output, and the help of equation (5.16), a *CI* for $\rho^2_{y.A}$ can be determined with boundaries $\rho^2_{y.A}(L)$ and $\rho^2_{y.A}(U)$.

This SAS-program **(VI_R2)** can be downloaded from this manual's webpage.

Example 5.6:

The unbiased estimator for $\rho^2_{y.A,B}$ in Example 5.5 is:

$$\begin{aligned} R_a^2 &= 0,3768 - (1 - 0,3768)\frac{3}{465 - 3 - 1} \\ &= 0,3768 - 0,0041 \\ &= 0,3727, \end{aligned}$$

while the 95% *CI* for $\rho^2_{y.A,B}$, is (0,308; 0,434) is and the input to the computer software package is: $u = 3,\ v = 461,\ F(3,461) = \dfrac{0,3768/3}{0,6232/461} = 92,9$ and $\alpha = 0,05$.

Since the sample was large enough, the R_a^2's value was, for practical purposes, the same as $R^2_{y.A}$. Further, this proportion variance is a large effect because even the lower bound of the *CI* lies comfortably above 0,25.

5.2.6 Confidence intervals for partial ρ^2

From Smithson (2001) and Cohen (1969, 1977, 1988) it follows that the sets A and B consist of *w* and *u* predictors respectively. The F-statistic for the partial R^2 is then given by:

$$F(u,v) = \frac{R^2_{yB.A}/u}{\left(1 - R^2_{yB.A}\right)/v}, \tag{5.17}$$

with $v = n - w - u - 1$.

Under the normality assumption, the statistic $F(u,v)$ follows a non-central F-distribution with non-centrality parameter:

$$ncp = \frac{\rho^2_{y.B.A}}{1-\rho^2_{yB.A}}\left(u+v+1\right) \ . \tag{5.18}$$

This is because the form of both $F\left(u,v\right)$ and ncp are the same as the statistics in equations (5.15) and (5.16). As before, computer packages can be used. The computer software program **VI_R2**, and the SAS-program **VI_R2pars** are then used with inputs *F*, *u*, *w* and *n*.

Example 5.7:
From the results of Example 5.5 it follows that $u = 2, w = 1,\ v = 465 - 1 - 2 - 1 = 461$ and

$$\mathrm{F}\left(2;461\right) = \frac{0,2869/2}{(1-0,2869)/461} = 92,74 \ .$$

With these values as inputs, Smithson (2001) constructed the following 90% *CI* for $\rho^2_{y.B.A}$: (0,2298; 0,3376). This interval indicates a large effect.

5.3 Effect sizes of the relationship between a continuous and a dichotomous variable

If a linear relationship between a continuous variable and a dichotomous variable has to be determined, the Pearson Product moment correlation coefficient can be calculated, which, in this situation, corresponds to the *point-biserial correlation*, denoted by r_{pb}. The population analogue is denoted by ρ_{pb}. As in paragraph 5.1, r_{pb} and ρ_{pb} can be used as effect size indices. Guideline values similar to those used for r_{xy} and ρ_{xy} are also applicable here. Cohen (1969, 1977, 1988) notes that if x and y are bivariate normally distributed with correlation ρ_{xy} and if x can be made into a dichotomous variable by splitting the values into two separate halves (concentrating on the median values of each half, with point-biserial correlation ρ_{pb}) then:

$$\rho_{xy} = 1{,}253\rho_{pb}. \tag{5.19}$$

This means that ρ_{pb} is almost 25% smaller than ρ_{xy} using the same data that was dichotomized. Consequently, Cohen's guideline values of 0,1; 0,3 and 0,5 can be similarly adapted.

5.3.1 Relationship between a continuous variable and group membership of two groups

If y has measurements on an interval/ratio scale (e.g., IQ, blood pressure, etc.) and x assumes only 2 values (e.g.,1 or 2) depending on the group membership of the y-variable measurement. Then, in terms of δ (described in equation in (4.2) (Cohen, 1969, 1977, 1988)) we have that:

$$\rho_{pb} = \frac{\delta}{\sqrt{\delta^2 + 1/(pq)}}, \tag{5.20}$$

where *p* is the proportion of elements (e.g., people) from the total number of elements in both populations that belong to the first population, while $q = 1 - p$ represents the remaining proportion of elements.

In the case where the two populations have equal numbers of elements, it follows from (5.20) that the point-biserial correlation is:

$$\rho_{pb} = \frac{\delta}{\sqrt{\delta^2 + 4}},$$

hence

$$\frac{1}{2}\delta = \frac{\rho_{pb}}{\sqrt{1 - \rho_{pb}^2}}. \tag{5.21}$$

The relationship between the probability of misclassification P_{MK} en ρ_{pb} follows from (4.39) en (5.21):

$$P_{MC} = 1 - U_2 = P\left(Z > \frac{\rho_{pb}}{\sqrt{1-\rho_{pb}^2}} \right). \tag{5.22}$$

The proportion misclassification is given in (4.39) as a function of δ and in (5.22) as a function of ρ_{pb}.

Formulas (5.20) and (5.21) allow us to convert δ-values to ρ_{xy}-values and vice versa.

Table 5.3 expresses the values of δ in terms of ρ_{pb} and ρ_{xy} , and provides guideline values for the former when we have $p = q = \frac{1}{2}$.

Table 5.3

Effect size	δ	ρ_{pb}	ρ_{xy}	Guideline values ρ_{xy}
Small	0,2	0,100	0,125	0,1
Medium	0,5	0,243	0,304	0,3
Large	0,8	0,371	0,465	0,5

It is clear that the recommended guideline values for ρ_{xy} naturally agree with the corresponding guideline values of δ as suggested by Cohen (see paragraph 4.5). Cohen's motivation concerning his choices of "small", "medium" and "large" in paragraph 4.5 serves to complement the motivation stated in paragraph 5.12.

An estimator for ρ_{pb} can be obtained from a random sample using the following expression provided by Kline (2004a:115):

$$\hat{\rho}_{pb} = \frac{\hat{\delta}}{\sqrt{\left(\hat{\delta}^2 + (n_1 + n_2 - 2)\cdot\frac{n_1 + n_2}{n_1 n_2}\right)}} \tag{5.23}$$

where $\hat{\delta}$ is the estimator of δ from (4.3).

In terms of the t-value of the t-test for two independent samples becomes:

$$\hat{\rho}_{pb} = \frac{t}{\sqrt{t^2 + n_1 + n_2 - 2}} \tag{5.24}$$

If one makes use of large samples of roughly the same sizes and variances, then P_{MC} is approximately the p-value of $\frac{\hat{\rho}_{pb}}{\sqrt{1-\hat{\rho}_{pb}^2}}$, i.e. :

$$\hat{P}_{MC} = 1 - \hat{U}_2 = P\left(Z > \frac{1}{2}d\right) = P\left(Z > \frac{\hat{\rho}_{pb}}{\sqrt{1-\hat{\rho}_{pb}^2}}\right). \tag{5.25}$$

Ozer (1985) provides an alternative method to estimate P_{MC} for equally sized samples:

If the dependent variable is dichotomised by dividing it into two equal halves (above and below the median) then the following 2 x 2 frequency table can be created:

	Below Me	Above Me	
Population A	a	n-a	n
B	n-a	a	n
	n	n	2n

The proportion of misclassifications of the dichotomised dependent variable is:

$$P_{MC}^1 = \frac{(n-a)+(n-a)}{2n} = \frac{n-a}{n} = 1 - \frac{a}{n}. \tag{5.26}$$

The phi-coefficient for the 2 x 2 – frequency table (see (5.41)) is, in this case:

$$\varphi = \frac{a^2 - (n-a)^2}{\sqrt{n.n.n.n}} = \frac{2na - n^2}{n^2} = \frac{2a}{n} - 1. \tag{5.27}$$

Now, since $\frac{1}{2}(1-\varphi) = \frac{1}{2}\left[1 - \frac{2a}{n} + 1\right] = 1 - \frac{a}{n}$, we find that

$$P_{MC} = \frac{1}{2}(1-\varphi). \tag{5.28}$$

By assuming that $\varphi = \rho_{pb}$, it follows that P_{MC} can also be determined by

$$P_{MC} = \frac{1}{2}(1-\rho_{pb}). \tag{5.29}$$

and estimated by

$$\tilde{P}_{MC} = \frac{1}{2}(1-\hat{\rho}_{pb}) . \tag{5.30}$$

An extract from Table 2 of Ozer (1985) illustrates how well this approximation performs:

δ	ρ_{pb}	P_{MC} (from (5.17b))	$\frac{1}{2}(1-\rho_{pb})$
0	0	0,5	0,5
0,2	0,1	0,46	0,45
0,63	0,3	0,38	0,35
1,15	0,5	0,28	0,25
1,96	0,7	0,16	0,15
4,13	0,9	0,02	0,05
∞	1,0	0	0

Counternull values for point-biserial correlation (see Rosenthal et.al, 2000: 15):

Equation (5.18) states $\hat{\delta}$ in terms of a point-biserial correlation $\hat{\rho}_{pb}$ if $n_1 = n_2$:

$$\hat{\rho}_{pb} = \frac{\hat{\delta}}{\sqrt{\hat{\delta}^2 + 4}}, \tag{5.31}$$

so that the counternull values of $\hat{\rho}_{pb}$ are approximately:

$$\hat{\rho}_{pb,counternull} = \frac{2\hat{\delta}}{\sqrt{4\hat{\delta}^2 + 4}} = \frac{\hat{\delta}}{\sqrt{\hat{\delta}^2 + 1}}. \tag{5.32}$$

From (5.19g) it follows that

$$\hat{\delta}^2 = \frac{4\hat{\rho}_{pb}^2}{1 - \hat{\rho}_{pb}^2}, \tag{5.33}$$

If we substitute (5.33) in (5.32) we find that:

$$\begin{aligned}\hat{\rho}_{pb,counternull} &= \sqrt{\frac{4\hat{\rho}_{pb}^2}{1 - \hat{\rho}_{pb}^2}} \Bigg/ \sqrt{\frac{4\hat{\rho}_{pb}^2}{1 - \hat{\rho}_{pb}^2} + 1} \\ &= \frac{2\hat{\rho}_{pb}^2}{\sqrt{1 + 3\hat{\rho}_{pb}^2}}.\end{aligned} \tag{5.34}$$

Example 7b:

For $\hat{\rho}_{pb} = 0,4$,

$$\hat{\rho}_{pb,counternull} = \frac{2 \times 0,4}{\sqrt{1 + 3 \times 0,4^2}} = 0,66 .$$

This means that this point-biserial correlation has the same probability of being as large as 0,66 as it does of being 0.

5.3.2 Modification for reliability

From paragraph 5.1.4 it follows from (5.7) that $\rho_{\rho b}$ and $\hat{\rho}_{\rho b}$ can be modified to obtain the reliability of y (which is denoted by ρ_{yy} or r_{yy}) as follows:

$$\rho_b = \frac{\rho_{pb}}{\sqrt{\rho_{yy}}} \text{ and } \hat{\rho}_b = \frac{\hat{\rho}_{pb}}{\sqrt{r_{yy}}} \quad . \tag{5.35}$$

From (5.21) and (5.23), the values δ and $\hat{\delta}$ can also be modified (see Baugh, 2003 : 36-38):

$$\delta_b = \frac{\rho_b}{\sqrt{1 - \rho_b^2}\sqrt{pq}} \tag{5.36}$$

and

$$\hat{\delta}_b = \frac{\hat{\rho}_b \sqrt{n_1 n_2}}{\sqrt{1-\hat{\rho}_b^2}\sqrt{(n_1+n_2-2)(n_1+n_2)}} \qquad (5.37)$$

Example 5.8:

From Example B in Chapter 3, we assume that the SDs for students and lecturers on the E/I measurements are both equal to $\sigma = 25$. Also, according to Rothmann et.al. (2000b), the reliability varies between $0,84$ and $0,86$, so that

$$\delta = \frac{94,58-107,64}{25} = -0,522$$

and

$$\rho_{pb} = \frac{\delta}{\sqrt{\delta^2 + 1/pq}} = \frac{-0,522}{\sqrt{0,522^2 + 1/(0,9\times 0,1)}} = -0,155$$

(where $p = \frac{254}{282} = 0,9$ and $q = 1-0,9 = 0,1$). Let $\rho_{yy} = 0,84$.

Now we have that $\rho_b = \frac{-0,155}{\sqrt{0,84}} = -0,169$.

$$\delta_b = \frac{-0,169}{\sqrt{1-0,169^2}\sqrt{0,9\times 0,1}} = -0,572$$

5.3.3 Proportion variance attributed to the group membership of two populations

While ρ_{pb} and $\hat{\rho}_{pb}$ can be used as effect size indices, in practice it is more sensible to make use of the square of this value. This squared quantity is then the proportion variance of y which can be attributed to the population group membership. Instead of using ρ_{pb}^2 and $\hat{\rho}_{pb}^2$, it is often even more useful to use the notation η^2 and $\hat{\eta}^2$. The reason for the use of η^2 is because, in the case of k populations, we defined it as $\eta^2 = \sigma_\mu^2 / \sigma_{tot}^2$, where σ_μ^2 is the variance of the values $\mu_1, \mu_2, ..., \mu_k$ and σ_{tot}^2 is the variance of all k populations together. In the case of two equal sized populations with equal variances, we obtain the expressions $\sigma_\mu^2 = \frac{1}{4}(\mu_1 - \mu_2)^2$ and $\sigma_{tot}^2 = \sigma_\mu^2 + \sigma^2$, so that from (5.20) with $p = q = \frac{1}{2}$ it follows that:

$$\eta^2 = \frac{\delta^2}{\delta^2 + 4} = \rho_{pb}^2 . \tag{5.38}$$

The value $\hat{\rho}_{pb}^2$ in (5.23) and (5.24) is actually a biased estimator of $\eta^2 \left(\rho_{pb}^2\right)$, and so, to compensate for this, Hays (see Sheskin, 2000: 264) proposes the quantity known as omega-squared:

$$\hat{\omega}^2 = \frac{t^2 - 1}{t^2 + n_1 + n_2 - 1} . \tag{5.39}$$

In terms of $\hat{\delta}$ this estimator becomes:

$$\hat{\omega}^2 = \frac{\frac{n_1 n_2}{n_1 + n_2}\hat{\delta}^2 - 1}{\frac{n_1 n_2}{n_1 + n_2}\hat{\delta}^2 + n_1 + n_2 - 1} . \tag{5.40}$$

The problem with $\hat{\omega}^2$ as an estimator is that it can be negative if |t| < 1. However, seeing as η^2 is positive (per definition), the effect size in these case is usually assumed to be zero. Cases where |t| < 1 are always associated with a non-statistically significant difference in group means and we expect that η^2 will be small.

Example 5.9:

Consider Example B from Chapter 3. Is there a practically significant difference between the mean preference score in E/I between students and lecturers? Assume equal standard deviations of $\sigma = 25,$ then

$\delta = (94,58 - 107,64)/25 = -0,522,$ and

$\eta^2 = \delta^2 / \left(\delta^2 + 1/pq\right),$ where $p = 254/282 = 0,9$ and $\mathrm{q} = 1 - 0,9 = 0,1$, so that we have

$$\begin{aligned}\eta^2 &= 0,522^2 / \left(0,522^2 + 1/(0,9 \times 0,1)\right)\\ &= 0,272/11,384\\ &= 0,024.\end{aligned}$$

The proportion variance of the E/I preference scores which is attributed to the two groups is only $0,024$.

Note that because the population sizes are very different it has a large influence on the value of η^2. In the case where the populations are equally large, $\left(p = q = \frac{1}{2}\right)$, then, for example:

$$\eta^2 = \delta^2 / \left(\delta^2 + 4\right) = 0,272/4,272 = 0,064$$

□

5.3.4 Guideline values for proportion variance attributed to population group membership

The guideline values suggested by Cohen (1969, 1977, 1988) for standardized differences, i.e., δ, is given in paragraph 4.5. Further, for correlations between two continuous variables, Cohen's guideline values given in paragraph 5.1.2 can be used. While in multiple linear regression, on the other hand, the guidelines are given in paragraph 5.2.4 for, among others, $\rho^2_{y.A}$. Proportion variance can be attributed to the population group membership and so, given ρ^2_{pb} and its estimator $\hat{\rho}_{pb}^2$, defined in terms of δ and $\hat{\delta}$ (see (5.17) and (5.18)). The guidelines for δ and $\hat{\delta}$ can also be used for ρ_{pb}^2 and $\hat{\rho}_{pb}^2$. Using Table 5.3, we get:

- Small effect: $\rho_{pb}^2 = 0,01$ $(\delta = 0,2 ; \rho_{pb} = 0,1)$
- Medium effect: $\rho_{pb}^2 = 0,06$ $(\delta = 0,5; \rho_{pb} = 0,243)$
- Large effect: $\rho_{pb}^2 = 0,14$ $(\delta = 0,8; \rho_{pb} = 0,371)$.

Notes:

1. These are in fact the guideline values that Cohen suggests for η^2, but in the case where there are more than two populations. In Chapter 6 we will discuss Cohen's motivation for these choices. In the present case ρ_{pb}^2 is actually not determined only through δ^2, but is also determined through the proportion p of population elements which belong to one of the populations (and q is then calculated from p). Table 5.4 provides values for ρ_{pb}^2 at selected values of δ and p.
2. From Table 5.4 it is clear that *as p becomes smaller, ρ_{pb}^2 also becomes smaller*. In the extreme case where $p = 0,01$, then ρ_{pb}^2 remains small or medium, according to the abovementioned guidelines. (Example 5.9 illustrates this nicely). Since $pq = p(1-p)$ is symmetric in p around the value $p = 0,5$, the same values of ρ_{pb}^2 are obtained for $p = 0,99;\ 0,95; \ldots 0,4.$

Table 5.4: Values of ρ_{pb}^2

δ \ p	0.01	0.05	0.1	0.15	0.2	0.3	0.4	0.5
0.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.2	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01
0.3	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.02
0.4	0.00	0.01	0.01	0.02	0.02	0.03	0.04	0.04
0.5	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.06
0.6	0.00	0.02	0.03	0.04	0.05	0.07	0.08	0.08
0.7	0.00	0.02	0.04	0.06	0.07	0.09	0.11	0.11
0.8	0.01	0.03	0.05	0.08	0.09	0.12	0.13	0.14
0.9	0.01	0.04	0.07	0.09	0.11	0.15	0.16	0.17
1	0.01	0.05	0.08	0.11	0.14	0.17	0.19	0.20
1.1	0.01	0.05	0.10	0.13	0.16	0.20	0.23	0.23
1.2	0.01	0.06	0.11	0.16	0.19	0.23	0.26	0.26
1.3	0.02	0.07	0.13	0.18	0.21	0.26	0.29	0.30
1.4	0.02	0.09	0.15	0.20	0.24	0.29	0.32	0.33
1.5	0.02	0.10	0.17	0.22	0.26	0.32	0.35	0.36
1.6	0.02	0.11	0.19	0.25	0.29	0.35	0.38	0.39
1.7	0.03	0.12	0.21	0.27	0.32	0.38	0.41	0.42
1.8	0.03	0.13	0.23	0.29	0.34	0.40	0.44	0.45
1.9	0.03	0.15	0.25	0.32	0.37	0.43	0.46	0.47
2	0.04	0.16	0.26	0.34	0.39	0.46	0.49	0.50

3. The case where $p = 0{,}5$ agrees with the above guideline values, is where it is assumed that the *populations are of equal size*.
4. Further, because ρ_{pb}^2 is a function of δ, the assumption of equal *standard deviations* of the two populations has to be made. This is because this assumption must be made in Example 5.9.
5. Grissom and Kim (2005: 92-95) give three reasons why the usage of ρ_{pb} is preferable to that of ρ_{pb}^2 .
6. Due to the restrictive assumptions made in 3 and 4 and what is mentioned in 5 above, it is recommended that, in the comparison of two population means, one should rather make use of one of the following effect size indices $\delta, \delta_a, \Delta, \Delta_1, \Delta_2, \delta_D$ and δ'_D or one of their estimators. One can choose an appropriate index which will satisfy the

assumptions of the situation. For this reason we will not spend further effort on the discussion of, for example, confidence intervals for ρ_{pb}^2.

5.4 Effect sizes for 2 x 2 - frequency tables

When population or sample elements can be classified using to two dichotomous categorical variables, this data can be expressed in a 2 $\times$ 2 – frequency or contingency table (also called a two-way table) as illustrated in Table 5.5 (see Steyn, 2002 and Kline, 2004a: 146):

Table 5.5

The 2 x 2 frequency table of x and y

		y Category 1	Category 2	Total
x	Category 1	a	b	$a+b$
	Category 2	c	d	$c+d$
	Total	$a+c$	$b+d$	n

In this table, a, b, c and d represent the frequencies at each of the 4 combinations of x and y's categories and $n=a+b+c+d$ represents the population or sample size (depending on whether we have sample or population data).

5.4.1 Relationships between x and y

The Pearson correlation coefficient between x and y (where each one can assume two values, e.g., 1 and 2) can be expressed in terms of the frequencies in Table 5.5 as follows:

$$\varphi = \frac{ad\text{-}bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}, \qquad (5.41)$$

which is known as the *phi-coefficient*. This coefficient has the same properties as ρ_{xy} and r_{xy} and, as such, can be used as an effect size index. Like ρ_{xy} and r_{xy}, the coefficient φ can also be negative, which will be the case when $bc>ad$. The order of the categories 1 and 2 is usually arbitrarily chosen, (e.g.,the first category of x is “male” and the second is “female”) and so, when possible, the frequency table can be constructed so that the larger frequencies at Category 1 of both x and y occur with Category 2 of both x and y. This is constructed so that the table produces a positive φ.

Cohen (1969, 1977, 1988) recommends, by referring to the guideline values for ρ_{xy}, the same values for φ, namely

- Small effect: $\varphi = 0{,}1$
- Medium effect: $\varphi = 0{,}3$
- Large effect: $\varphi = 0{,}5$.

5.4.2 Binomial Effect Size Display (Rosenthal et.al 2000: 17)

To interpret the φ- coefficient in terms of a 2 x 2 frequency table, the so-called BESD (“Binomial Effect Size Display”) is used. For example, consider an experimental group and a control group, each of size 100, and suppose that of the 200 individuals in the study 100 showed improvements after a treatment and that the 2 x 2 table of results was given as follows:

		Improvement	No improvement	Total
	Experimental	66	34	100
Groups				
	Control	34	66	100
	Total	100	100	200

We find that $\varphi = \dfrac{66\times 66 - 34\times 34}{\sqrt{100\times 100\times 100\times 100}} = 0,32.$

In general the content of the 2 x 2 table is given by:

$100\left(0.5+\dfrac{r}{2}\right)$	$100\left(0.5-\dfrac{r}{2}\right)$
$100\left(0.5-\dfrac{r}{2}\right)$	$100\left(0.5+\dfrac{r}{2}\right)$

where $r = \varphi$. In this example $r = 0{,}32$ so that 66% - 34% = 32%. It thus provides the difference in improvement rates (66% vs 34%) if half of the population receives the treatment (belonging to the experimental group) and the other half does not receive the treatment (belonging to the control group). The following table provides the improvement rates for values of $r = \varphi$:

Rate of improvement

$r = \varphi$	**From**	**To**	**Effect** (Cohen, 1988)
0,0	0,5	0,5	
0,1	0,45	0,55	small
0,2	0,40	0,60	
0,3	0,35	0,65	medium
0,4	0,30	0,70	
0,5	0,25	0,75	large
0,6	0,20	0,80	
0,7	0,15	0,85	
0,8	0,10	0,90	
0,9	0,05	0,95	
1,0	0,00	1,00	

In order to get an intuitive "feeling" for the φ-values in terms of frequency tables, Steyn (2002) provides the following examples, summarized in Table 5.6:

Table 5.6

Examples of 2 x 2 tables

(a) $\varphi = 0$: if the frequencies in 2 rows (or columns) are equal, e.g.,

		y 1	2	
x	1	50	50	100
	2	25	25	50
		75	75	150

(b) $\varphi = 0,1$: (small effect):

		y 1	2	
x	1	55	45	100
	2	45	55	100
		100	100	200

(c) $\varphi = 0,3$: (medium effect):

		y 1	2	
x	1	65	35	100
	2	35	65	100
		100	100	200

(d) $\varphi = 0,5$: (large effect):

		y 1	2	
x	1	75	25	100
	2	25	75	100
		100	100	200

(e) $|\varphi| = 1$: if the frequencies in any diagonal entry of the table is equal to 0, e.g.,

		y 1	2	
x	1	100	0	100
	2	0	100	100
		100	100	200

Table 5.6(e) is an example of a *strict perfect relationship* between x and y (Smithson, 2000: 324). This means that x completely determines y and vice

versa. If an individual obtains a 1 for x, that person will also obtain a 1 for y, while everyone with a 2 for x will also obtain a 2 for y.

Now consider the following table:

		y 1	2	
x	1	100	0	100
	2	75	25	100
		175	25	200

This is an example a weak *perfect relationship* (Smithson, 2000: 324) in the sense that only category 1 of x will completely determine the y value, but category 2 cannot be used to determine y at all. Alternatively, x can be fully determined if $y=2$. In this case $\varphi = 0{,}38$ which is a considerable decrease from $\varphi = 1$. This indicates that φ *is not an appropriate measure a weak perfect relationships.* We will show later that the "Odds ratio" is more appropriate for this purpose.

Example 5.10:

In Example C, Chapter 3, combines the last 3 Categories of smoking together, so that it forms a 2 x 2 - table:

Coronary heart disease

		Yes	No	Total
Smoke	Yes	78	59	137
	No	42	61	103
	Total	120	120	240

With the goal of determining the relationship between coronary heart disease and smoking, the coefficient φ is calculated as follows

$$\varphi = \frac{78 \times 61 - 59 \times 42}{\sqrt{137 \times 103 \times 120 \times 120}} = \frac{2280}{\sqrt{203198400}} = 0{,}16,$$

which indicates a small effect.

Suppose that the 240 employees are randomly chosen from all possible employees at a company. In this case φ would be estimated by the value $0{,}16$.

□

In general, the sample value of φ, denoted $\hat{\varphi}$, can be used as an estimator of the population value of φ. This estimator is asymptotically unbiased, but overestimates φ for small samples by approximately ${}^{1}\!/\!_{\sqrt{n}}$ (Johnson et.al, 1995: 447).

Note:

Fleiss (1994) demonstrates the following problem with φ as an effect size index, by making use of the following example. Consider two studies where the relative frequencies for y for a given x is the same, but the relative frequencies of x differs:

			y		
Study			+	-	Total
1	x	+	45	5	50
		-	120	30	150
	Total		165	35	200
2	x	+	90	10	100
		-	80	20	100
	Total		170	30	200

In both studies for x (+) the relative frequencies are 45/50 = 90/100 and 5/50 = 10/100, and similarly for x (-). The relative frequencies for x (+) are 50/200 and 100/200 which differ – similarly for x (-).

The φ - coefficient is 0,11 and 0,14 for the two studies.

This means that the φ - coefficient is influenced by the degree which the categories of x are represented in the data. The same is also true for the y - categories.

For this reason $\hat{\varphi}$ is a *valid estimator* **if** it is based on something other than a *random sample*. For randomness, the marginal totals of the 2 x 2 – frequency table should appear in roughly the same ratios as that of the population. Now, consider the following fictional frequency table obtained from Example 5.10, but where a random sample is drawn from the company instead of a stratified sample with an equal number of employees with and without heart disease:

		Coronary heart disease		
		Yes	No	Total
Smoke	Yes	26	98	124
	No	14	102	116
	Total	40	200	240

This table is obtained from the 240 employees which are divided into 40 people with the heart disease (instead of 120) and by taking the number of smokers as one third of the original 78. Similarly, $\frac{59}{120} \times 200$ is approximately 98 rounded off to the nearest integer. This table should be a realization of a random sample if one sixth (i.e., $\frac{40}{240}$) of the employees have heart disease. The value $\hat{\varphi} = 0,119$ is a valid estimator for the population φ - coefficient, while Example 5.10's value of $\hat{\varphi} = 0{,}16$, based on the stratified sample is not a valid estimator.

5.4.3 The counternull of the BESD

The φ-coefficient can be seen as a special case of a point-biserial correlation where the response variable (y) is dichotomous. We obtained a counternull value for the estimator $\hat{\rho}_{pb}$ in (15.19d) in terms of $r = \varphi$:

$$r_{counternull} = \frac{2r}{\sqrt{1+3r^2}}. \qquad (5.42)$$

For a BESD where $\varphi = r$, the counternull of a BESD is one with $\varphi = r_{counternull}$. In the above example with $\varphi = 0{,}32$ we have

$$r_{counternull} = \frac{2 \times 0{,}32}{\sqrt{1+3\times 0{,}32^2}} = 0{,}56,$$

So that the counternull BESD is:

Group	Improve	Not improve	Total
Experimental	78	22	100
Control	22	78	100
Total	100	100	200

This 2 x 2 – table is thus just as probable as the BESD where $\varphi = 0$, viz.:

Group	Improve	Not improve	Total
Experimental	50	50	100
Control	50	50	100
Total	100	100	200

5.4.4 Confidence interval for φ

For large samples Fleiss (1994) provides the variance of $\hat{\varphi}$ as:

$$Var(\hat{\varphi}) = \frac{1}{n}\left[1 - \hat{\varphi}^2 + \hat{\varphi}\left(1 + \frac{\hat{\varphi}^2}{2}\right)C_1 - \frac{3}{4}\hat{\varphi}^2 C_2\right], \qquad (5.43)$$

where

$$C_1 = \frac{(a+b-c-d)(a+c-b-d)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

and

$$C_2 = \frac{(a+b-c-d)^2}{(a+b)(c+d)} + \frac{(a+c-b-d)^2}{(a+c)(b+d)}$$

The approximate $100(1-\alpha)\%$ *CI* for φ's boundaries are thus:

$$\varphi_L = \hat{\varphi} - z_{\alpha/2}\sqrt{Var(\hat{\varphi})}$$

and (5.44)

$$\varphi_U = \hat{\varphi} + z_{\alpha/2}\sqrt{Var(\hat{\varphi})} \ .$$

As an alternative, the exact *CI* can be determined by making use of the SAS-program **VI_w** for φ as a special case of *w* discussed in paragraph 5.5.2. The inputs used are $X2 = n\hat{\varphi}^2$, *n* and *df*=1.

Example 5.11:

For Example 5.10

$$C_1 = \frac{(137-103)(120-120)}{\sqrt{137\times 103\times 120\times 120}} = 0$$

$$C_2 = \frac{(137-103)^2}{137\times 103} + 0 = 1156/14111 = 0{,}082$$

$$Var(\hat{\varphi}) = \frac{1}{240}\left[1 - 0{,}16^2 + 0{,}16\left(1 + \frac{0{,}16^2}{2}\right)\times 0 - \frac{3}{4}\times 0{,}16^2 \times 0{,}082\right]$$

$$= \frac{1}{240} \times 0,9728 = 0,00405$$

The 95% *CI*'s boundaries are then:

$$\varphi_L = 0,16 - 1,96\sqrt{0,00405} = 0,16 - 0,125 = 0,035$$

$$\varphi_U = 0,16 + 0,125 = 0,285.$$

For the exact *CI* is the input $X2 = 240 \times (0,16)^2 = 6,144$, *n*=240 and *df*=1.

This produces the 95% *CI* of (0,032 ; 0,287), which is very close to the approximate *CI* value.

Thus, even with a large sample (like 240) the 95% *CI*'s boundaries are quite wide and the φ's value vary in such a way that it indicates a small to medium effect. □

5.4.5 Probability measures from 2 x 2 frequency tables

Suppose that the proportion of population elements in populations 1 and 2 are p and q respectively. Suppose that the response value (y) can only take on either a positive or negative value (e.g., 'agree' versus 'disagree'; in case control studies in epidemiology 'exposed' versus 'not exposed'; in intervention studies 'improves' versus 'does not improve'). Let the probabilities (proportions) for positive responses be π_1 and π_2 for both populations. The 2 x 2 frequency table then has the following form:

Table 5.7 General 2 x 2 – table

		y Positive	Negative	Total
x	Population 1	$pN\pi_1$	$pN(1-\pi_1)$	pN
	2	$qN\pi_2$	$qN(1-\pi_2)$	qN
	Total	$N\pi$	$N(1-\pi)$	N

The expression $\pi = p\pi_1 + q\pi_2$ represents the probability of a positive response for both populations, while *N* is the total number of elements in both populations.

By making use of Table 5.7, we can discuss the following three comparative *measures of risk or rate:*

- Difference in proportion of the positive responses, $\pi_1 - \pi_2$.
- Ratio of the proportion of the positive responses, π_1 / π_2, known as the relative risk or rate or risk ratio.
- Ratio of the odds, or odds ratio

$$\omega = \frac{\pi_1 / (1 - \pi_1)}{\pi_2 / (1 - \pi_2)} = \frac{\pi_1 / (1 - \pi_2)}{\pi_2 / (1 - \pi_1)}.$$

Measures of rates is a more general term, because we tend to only use the term *measure of risk* when the 'positive' response represents some undesirable result such as 'exposed', 'identified', 'sick', or 'dead'.

5.4.6 Difference in proportions

As in the case of the mean, in this situation there are two types of effect size indices that can be used. First, the standardized differences in proportions can be used, and second, one can use the relationship between the response y and the population grouping variable, x.

(a) Standardized differences in proportions:

Let

$$y_i = \begin{cases} 1, & \text{if population } i \text{ is positive} \\ 0, & \text{otherwise,} \end{cases}$$

then y_i's population mean is $\mu_i = \pi_i$ and its population variance is $\sigma_i^2 = \pi_i(1 - \pi_i)$. This means that, from (4.21) with $W_1 = p$ and $W_2 = q$ it follows that:

$$\delta_g = \frac{\pi_1 - \pi_2}{\sqrt{p\pi_1(1-\pi_1) + q\pi_2(1-\pi_2)}} \quad . \tag{5.45}$$

If equal population variances are assumed, then each of the variances with $\pi(1-\pi)$ (recall that $\pi = p\pi_1 + q\pi_2$) can be replaced. This yields the proportion analogue of δ:

$$\delta = \frac{\pi_1 - \pi_2}{\sqrt{\pi(1-\pi)}} \tag{5.46}$$

Starting with population 1 as the reference point (say the control population), the effect size-index is:

$$\Delta_1 = (\pi_1 - \pi_2) / \sqrt{\pi_1(1-\pi_1)} \tag{5.47}$$

The estimators $\hat{\delta}_g$, $\hat{\delta}$ and Δ_1 can be obtained by replacing the proportions π_1 and π_2 with the sample proportions p_1 and p_2, for the two populations.

The problem with all three of the above indices is that the standard deviation value depends on π_1 and π_2.

Cohen (1969, 1977, 1988) therefore proposes the following *effect size index*:

$$\psi = 2\left(arc\,sin\left(\sqrt{\pi_1}\right) - arc\,sin\left(\sqrt{\pi_2}\right)\right) \tag{5.48}$$

Note that for $arc\,sin(x)$ the angle is expressed in radians (a), such that $sin(a) = x.$

Notes:

- The function $arc\,sin(x)$ is also often expressed as $sin^{-1}(x)$ on calculators.
- Radians can be converted to degrees using the relationship $a = \frac{\theta}{360} \times 6,283,$ where θ is the angle in degrees.

- If $\pi_1 = 0$ or $\pi_2 = 0$, then let $arc\,sin\left(\sqrt{1/(4n)}\right)$ instead of $arc\,sin(0)$.
- If π_1 or $\pi_2 = 1$, use 1,571 - $bg\,sin\left(\sqrt{1/(4n)}\right)$ instead of $arc\,sin(1)$.
- The standard deviation of ψ is independent of π_1 and π_2, so that by comparison of means the scale remains constant. For example, for $\pi_1 = 0,65$ and $\pi_2 = 0,35$ the value ψ is $\psi = 0,61$, while for $\pi_1 = 0,5$ and $\pi_2 = 0,2$ it is $\psi = 0,64$. This means that a difference of $0,3$ in proportions roughly equates into a difference of $0,6$ on the ψ-scale. For the index δ_g, the corresponding values would have been $0,63$ and $0,50$, if it was assumed that $p = q = \frac{1}{2}$.
- Note that for a BESD-2x2-table (see previous paragraph), all the marginal totals are 100 and $\pi_1 - \pi_2$ has the value of φ. Therefore r = $\pi_1 - \pi_2$, so that the BESD can be determined from the difference in the proportions.

If random samples are drawn from the populations, the sample estimates p_1 and p_2 will serve as estimators for the proportions, and the following estimator can be used:

$$\hat{\psi} = 2\left(arc\,sin\left(\sqrt{p_1}\right) - arc\,sin\left(\sqrt{p_2}\right)\right) \tag{5.49}$$

For large samples, $\hat{\psi}$ is normally distributed with mean ψ and variance $\left(1/n_1 + 1/n_2\right) = \frac{n_1 + n_2}{n_1 n_2}$. This means that boundaries of the $100(1-\alpha)\%$ *CI* can be expressed as:

$$\psi_L = \hat{\psi} - z_{\alpha/2}\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

and (5.50)

$$\psi_U = \hat{\psi} + z_{\alpha/2}\sqrt{\frac{n_1+n_2}{n_1 n_2}}.$$

Counternull value of $\hat{\psi}$:

Since $\hat{\psi}$ is normally distributed with variance independent of ψ, it follows from paragraph 4.1 that, in the same way as in the case of d, die counternull value for $\hat{\psi}$ is given by $2\hat{\psi}$.

Example 5.12:

Consider Example 5.10 and denote the population of coronary heart disease sufferers as population 1 and the population without the disease as population 2. Then $\pi_1 = \frac{78}{120} = 0,65$ $\pi_2 = 0,49$, $p = \frac{120}{240} = 0,5 = q.$

$$\begin{aligned}\delta_g &= \frac{0,65 - 0,49}{\sqrt{0,5\times 0,65\times 0,35 + 0,5\times 0,49\times 0,51}}\\ &= \frac{0,17}{\sqrt{0,1138+0,1250}} = \frac{0,17}{0,489} = 0,348.\end{aligned}$$

To determine δ, we calculate $\pi = \frac{137}{240} = 0,57$ so that $\delta = 0,17/\sqrt{0,57\times 0,43} = 0,17/0,495 = 0,343$, which, for all practical purposes, is the same as δ_g.

$$\begin{aligned}\psi &= 2\left(arc\,sin\left(\sqrt{0,65}\right) - arc\,sin\left(\sqrt{0,49}\right)\right)\\ &= 2\left(0,9377 - 0,7754\right) = 0,325,\end{aligned}$$

which produces almost the same effect size.

If we assume that two random samples are drawn from populations 1 and 2, then an approximate 95% *CI* for ψ ‘s has the following boundaries:

$$\psi_L = 0,325 - 1,96\sqrt{\frac{120+120}{120\times120}}$$
$$=0,325 - 1,96\times0,129 = 0,072$$
$$\psi_U = 0,325 + 0,253 = 0,578$$

The population value ψ can thus be as low as $0,072$ but also as high as $0,578$ (with probability 95%).

5.4.7 Guideline values for differences in proportions

From Example 5.12 it seems as though all three effect size-indices δ_g, δ and ψ produce almost identical values. This is true in practice for all combinations of the values:

$$0,1 \le \pi_1, \pi_2 \le 0,9 \quad \text{and} \quad 0,25 \le p \le 0,5 .$$

Using the guidelines based on the mean discussed for δ, Cohen (1969, 1977, 1988) once again proposes the same guidelines:

Small effect: $\delta,\ \delta_g,\ \psi = 0,2$. These guideline values are used if $(\pi_1;\pi_2)$ forms the following pairs: $(0,005;0,1),(0,2;0,29),(0,4;0,5),(0,6;0,7),(0,8;0,87)$ and $(0,9;0,95)$.

Medium effect: $\delta,\ \delta_g,\ \psi = 0,5$. These values are used if (π_1,π_2) form the pairs: $(0,05;0,21),(0,2;0,43),(0,4;0,65),(0,6;0,82),(0,8;0,96)$.

Large effect: $\delta,\ \delta_g,\ \psi = 0,8$. These values are used if (π_1,π_2) form the pairs: $(0,05;0,34),(0,2;0,58),(0,4;0,78),(0,6;0,92),(0,8;0,996)$.

Burnand et. al.(1990) proposes the following guidelines which were determined empirically from a survey of 392 articles in the medical literature:

- Significant : δ = 0,28
- Substantially significant : δ = 0,35

- Highly significant : δ = 0,65.

5.4.8 Rate or Risk ratios

The rate ratio is the ratio of the probabilities π_1 and π_2 as defined in paragraph 5.4.5. If population 1 is the control population and population 2 is the treatment population, then π_1 / π_2 is the ratio of the proportion of positive response of the control individuals relative to the treated individuals. If the term 'positive' refers to the event of occurrence of a disease or death, then it is known as a risk ratio. If $\pi_1 / \pi_2 > 1$, then it means that the risk is larger for the control group than it is for the treatment group. If definitions of population 1 and 2 are reversed, then $\pi_1 / \pi_2 < 1$ would be more beneficial.

The calculation of π_1 / π_2 in terms of the cell-frequencies of a 2 x 2- frequency table (Table 5.5) is:

$$\pi_1 / \pi_2 = \frac{a / (a+b)}{c / (c+d)} \tag{5.51}$$

If one is working with sample data, then the estimated rate ratio is p_1 / p_2, where p_1 and p_2 are the sample proportions based on samples drawn from the two populations.

One disadvantage of the rate ratio is that it can become very large if π_2 becomes very small relative to π_1.

For this reason it cannot be used as an effect size index in the same manner as φ or η^2, both of which lie between 0 and 1. However, it should rather be used by looking how far it lies from 1, because $\pi_1 / \pi_2 = 1$ implies that there is no difference between the rates or risks. The natural logarithm of π_1 / π_2, i.e., $\ell n(\pi_1 / \pi_2) = \ell n(\pi_1) - \ell n(\pi_2)$ will serve as a more effect size-index, since it can assume any value on the number line and the zero value corresponds to no differences in the rates or risks.

According to Fleiss (1994) and Kline (2004a), the natural logarithm $\ell n(p_1 / p_2)$ follows as approximate normal distribution if the sample is large. Further, we find that

$$Var\left[\ell n(p_1 / p_2)\right] = \frac{1-p_1}{n_1 p_1} + \frac{1-p_2}{n_1 p_2}, \tag{5.52}$$

so that the $100(1-\alpha)\%$ *CI* for $\ell n(\pi_1 / \pi_2)$ has lower and upper boundaries given by the following expression:

$$\ell n(p_1 / p_2) \pm z_{\alpha/2}\sqrt{\frac{1-p_1}{n_1 p_1} + \frac{1-p_2}{n_2 p_2}} \tag{5.53}$$

The *CI* for the quantity π_1 / π_2 then has the following lower and upper boundaries:

$$(\pi_1\pi_2)_{\mathrm{L}} = e^{\mathrm{L}} \text{ and } (\pi_1\pi_2)_U = e^U \tag{5.54}$$

Example 5.13:

Continuing with Example 5.6, let population 1 denote individuals with coronary heart disease and population 2 denote individuals without the disease, then the probability that an individual from the respective populations smoke are $\pi_1 = \frac{78}{120} = 0,65$ and $\pi_2 = \frac{59}{120} = 0,49$.

Consequently, $\pi_1 / \pi_2 = 0,65/0,49 = 1,327$, which means that an individual with coronary heart disease is 1,3 more likely to smoke than an individual without the disease. Smoking might thus be seen as risk factor for the disease.

If the 120 individuals in each group are considered to be random samples, then $p_1 = 0,65$ and $p_2 = 0,49$ and $p_1 / p_2 = 1,327$ are estimates of the rate ratio, while

$$\ell n(p_1 / p_2) = \ell n(1,327) = 0,283$$

$$Var\left[\ell n\left(p_1 / p_2\right)\right] = \frac{1\text{-}0,65}{120 \times 0,65} + \frac{1\text{-}0,49}{120 \times 0,49}$$
$$= 0,00449 + 0,00864$$
$$= 0,0131$$

the 95% *CI* for $\ell n\left(\pi_1 \pi_2\right)$ is then:

$$\begin{aligned} & 0,283 \pm 1,96\sqrt{0,0131} \\ &= 0,283 \pm 0,225 \\ &= \left(0,058; 0,508\right) \end{aligned}$$

$\left(\pi / \pi_2\right)_L = e^{0,058} = 1,060$, $\left(\pi_1 / \pi_2\right)_U = e^{0,508} = 1,661$.

This means that the quantity π_1 / π_2 can be as low as 1,06 and as high as 1,661 with 95% probability. There is thus a sign of a risk.

Odds ratio:

We will begin by first defining the odds. In terms of Table 5.7, the odds of population 1 is $\pi_1 / \left(1 - \pi_1\right)$ and for population 2 it is $\pi_2 / \left(1 - \pi_2\right)$. It thus denotes the ratio of the probability of y yielding a positive result versus probability of y yielding a negative result.

Example 5.14:
In Example 5.10, the odds of a person with coronary heart disease is

$$\frac{78}{120} \Big/ \frac{42}{120} = \frac{78}{42} = 1,857,$$

while, the odds of a person without the disease is

$$\frac{59}{120} \Big/ \frac{61}{120} = 0,967 .$$

For individuals with the heart disease, there are approximately 1,9 people who smoke for each one that does not smoke, while for the people without heart disease it is closer to 1. □

If the two populations' odds are to be compared, then it can be done by looking at the ratio of the odds.

This ratio is commonly known as the odds ratio, or simply OR. It is expressed as:

$$\omega = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)} = \frac{ad}{bc} \quad (5.55)$$

For calculation purposes, it is easiest to use the equation $\frac{ad}{bc}$ using the values defined in Table 5.5. The value of the OR can vary between 0 and infinity, with the value of 1 indicating that the two odds are equal. The values 0 and infinity are obtained if any of the frequencies in the 2 x 2 table are equal to 0. In paragraph 5.4.1 this case is described as a weak perfect relationship between x and y.

Example 5.15 (Smithson, 2000: 324):

A clinical psychologist has done research concerning snake phobias. The following table is obtained:

		Dislike snakes		
		No	Yes	Total
Fear of	Yes	5(b)	49(a)	54
Snakes	No	49(d)	159(c)	208
	Total	54	208	262

The odds of people with a fear of snakes = $49/5 = 9,8$.

The odds of people who do not have a fear of snakes = $159/49 = 3,24$.

The odds ratio is:

$$OR = 9{,}8/3{,}24 = 3{,}02\,.$$

(Note that OR can also be obtained as follows

$$OR = (\mathrm{ad})/(\mathrm{cd}) = (49 \times 49)/(5 \times 159) = 3{,}02\,.$$

(The frequency of the Yes-Yes category is *a*, etc.).

This means that the odds of a person with a fear for snakes is 3 times higher than for those people without a fear of snakes. An *OR* value of $\frac{1}{3{,}02} = 0{,}331$ will have the same interpretation if we compare the odds of a person without a fear of snakes with the odds of a person with a fear of snakes.

Smithson (2000: 326) states two reasons why the *OR* is preferable as a measure of the relationship to the φ-coefficient:

1) It serves as an effective measure of weak perfect relationships;
2) It remains the same even if a row or column in the 2 x 2 – table is multiplied by a factor.

If the random sample is drawn from a population, the population OR(ω) is estimated by $\hat{\omega}$, where a, b, c and d are the sample frequencies. If b or c or both are equal to zero, then $\hat{\omega}$ is undefined. The estimator proposed by Jewell (see Shoukri & Chaudhary, 2007) should then be used viz.

$$\hat{\omega}_J = \frac{\mathrm{ad}}{(\mathrm{b}+1)(\mathrm{c}+1)}$$

For Monte-Carlo simulations with n = 25 one finds that $\hat{\omega}_J$ has a smaller bias and mean squared error when compared to other estimators like $\hat{\omega}$.

Example 5.15 illustrates a near weak perfect relationship (the cell frequency of 5 is "close to" zero). In this case OR=3,02 and if the first row and column's frequencies were 3 and 51, then the value would change to $5,522$ and becomes infinitely large if the frequencies were 0 and 54. The φ-coefficient for Example 5.15, on the other hand, is 0,143 and increases to 0,26 if the first cell frequency is set to 0. This indicates that it is far from a perfect relationship. This helps to illustrate advantage 1) discussed above

As far as advantage 2) is concerned, we will refer to the note stated in paragraph 5.4.1 where two studies with different relative frequencies for x's two categories produce two different φ-values; i.e., $0,11$ and $0,14$. For these two studies the OR – values are actually the same: Study 1 : $\frac{45 \times 30}{120 \times 5} = 2,25$

Study 2 : $\frac{90 \times 20}{80 \times 10} = 2,25$.

Like the rate ratio, π_1 / π_2, the *OR* is usually judged by interpreting its distance from 1.

Therefore, the natural logarithm of *OR,* $\ell n(\omega)$, is sometimes easier to use because the distance from 1 on the original scale converts to a distance from 0 on the natural logarithm scale.

If a random sample is drawn from a population, then the population's *OR* (ω) is estimated by $\hat{\omega}$, where *a, b, c* and *d* are the sample frequencies. For large samples we find that $\ell n(\hat{\omega})$ follows an approximate normal distribution with mean $\ell n(\omega)$ and variance (Fleiss, 1994)

$$Var\left[\ell n(\hat{\omega})\right] = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \tag{5.56}$$

Therefore, a $100(1-\alpha)\%$ *CI* for $\ell n(\omega)$ has the following boundaries $(L;U)$

$$\ell n\left(\frac{ad}{bc}\right) \pm z_{\alpha/2}\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} \ . \tag{5.57}$$

As a result, the boundaries for the *CI* for ω are

$$\omega_{\mathrm{L}} = e^{\mathrm{L}} \quad \text{and} \quad \omega_U = e^U \ . \tag{5.58}$$

Further applications of odds ratios are discussed in Fleiss (1994):

- When other variables (covariates) related to the response variable y, apart from grouping variable x, are included, one can apply a logistic regression analysis, from which the *OR*-value can be obtained directly.
- The Mantel-Haenszel estimator is used whenever the covariates in the above situation stated above are categorical (i.e., the data is distributed into strata) and is another method to combine the values of $log(OR)$.

Newcombe (2006) provides the following reasons why OR is the most commonly used measure for 2 x 2 –frequency tables:

a. The OR’s natural role in logistic regression
b. It is the only sensible measure when one makes use of retrospective case-controlled study designs (as is often the case in epidemiological studies) instead of probability sampling.
c. When the occurrence of an event (a disease for example) is rare, the OR’s value is very similar to the risk ratio, since a / b $\approx$ a / (a+b) and c / d $\approx$ c / (c+d).

However, Newcombe warns that the OR value will always lie further away from 1 than the risk ratio value and will thus overestimate the risk. Additionally, OR is use as if it is identical to the risk ratio, which is only the case in rare circumstances.

Example 5.16:

If the frequency table in Example 5.15 is the result of a random sample, and the resulting estimated odds ratio is $\hat{\omega} = 3{,}02$ and the 90% *CI* for $\ell n(\omega)$ is:

$$\ell n(3,02) \pm 1,645\sqrt{\frac{1}{5}+\frac{1}{49}+\frac{1}{49}+\frac{1}{159}}$$
$$= 1,105 \pm 2,645 \times 0,497$$
$$= (0,2873;1,9227)$$

Thus, a 90% *CI* for ω is: $(1,333;6,840)$, and the population *OR* can be as small as $1,33$ and as large as $6,84$ with a 90% probability. □

5.4.9 Interpretation of *OR* as an effect size

According to Kline (2004a:147) and Chinn (2000) an *OR* can be reduced to a standardized difference analogously to δ. However, since $\ell n\left[\pi_1/(1-\pi_1)\right]$ and $\ell n\left[\pi_2/(1-\pi_2)\right]$ each follow a logistic distribution, which is approximately normal with standard deviation $\pi/\sqrt{3} = 1,81$, and the standardized difference is thus

$$\delta_{OR} = \frac{\ell n\left[\pi_1/(1-\pi_1)\right] - \ell n\left[\pi_2/(1-\pi_2)\right]}{1,81} = \frac{logit(\pi_1) - logit(\pi_2)}{1,81} = \frac{\ell n(\omega)}{1,81} \qquad (5.59)$$

The standardized difference δ_{OR} can be dealt with in a similar manner as δ, and the same guideline values can be used, i.e.,

Small effect : $\delta_{OR} = 0,2$

Medium effect : $\delta_{OR} = 0,5$

Large effect : $\delta_{OR} = 0,8$.

From (5.59) it follows that $\omega = e^{1,81\delta_{OR}}$, therefore:

Small effect : $\omega = 1,44$, which will be taken as $1,5$

Medium effect : $\omega = 2,48$. which will be taken as $2,5$

Large effect : $\omega = 4,27$, which will be taken as $4,25$

While an *OR* value larger than 1 indicates that the odds of the first population is indeed larger than the other, it does not necessarily indicate an important difference in the odds. In other words, while the values $0,5$ and $0,8$ for δ_{OR}, per the recommended guideline values, should be medium and large effects, the guideline values of $2,5$ and $4,25$ for the *OR* are suggested for medium and large effects.

Based on a survey of medical journals which consisted of 392 articles, Burnand et. al. (1990) suggest the following guideline values for OR:

- Significant : OR = 2,2
- Substantially significant : OR = 2,5
- Highly significant : OR = 4,0.

The last two guideline values correspond to 'medium' and 'large' effects.

Because $\hat{\delta}_{OR}$ has a large sample normal distribution with variance independent of $\hat{\delta}_{OR}$, then it follows, similarly to ψ, that the counternull values are $2\hat{\delta}_{OR}$. From (5.59) we find that

$$\hat{\omega}_{counternull} = e^{1,81\hat{\delta}_{OR-counternull}} = e^{3,62\hat{\delta}_{OR}} \quad . \qquad (5.60)$$

A further interpretation of OR is as follows:

According to Tritchler (1995) we find, for two normally distributed populations (pop.1 and pop.2) with means μ_1 and μ_2 and common standard deviation σ that

E = P(Classify x in pop.1 | x is from pop.2)

= P(Classify x in pop. 2 | x is from pop.1)

$$= \Phi\left(-\frac{1}{2}\delta\right), \qquad (5.61)$$

where Φ(t) is the cumulative distribution function of a standard normal distribution and

$$\delta = \frac{|\mu_1 - \mu_2|}{\sigma},$$

the standardised absolute difference in means as defined in Chapter 4.

The special case of the linear classification rule used in discriminant analysis (see Chapter 8) in univariate populations then simplifies to:

Classify x in pop.1 if

$$x > \frac{\mu_1 + \mu_2}{2} \quad \text{if} \quad \mu_1 > \mu_2 \,.$$

Tritchler (1995) then proposes the following joint probability of the two dichotomies $x > c, x \le c$ for Pop. 1 and Pop. 2:

	Pop. 1	Pop. 2
$x > c$	(1-E).P(1)	E.P(2)
$x \le c$	E.P(1)	(1-E).P(2)

where P(i) = P(x is from pop.1).

The odds ratio of this 2 x 2 table is thus:

$$\omega = \left(\frac{1-E}{E}\right) \Big/ \left(\frac{E}{1-E}\right) = \left(\frac{1-E}{E}\right)^2 , \qquad (5.62)$$

and from (5.43b) it follows that

$$\omega = \frac{1 - \Phi\left(-\frac{1}{2}\delta\right)}{\Phi\left(-\frac{1}{2}\delta\right)} . \qquad (5.63)$$

Using the guidelines in Cohen (1988), then from (5.63):

Small effect: $\delta = 0,2$: $\omega = 1,38$.

Medium effect: $\delta = 0,5$: $\omega = 2,25$.

Large effect: $\delta = 0,8$: $\omega = 3,64$.

These values of ω correspond relatively well with those obtained from δ_{OR} and those proposed by Burnand et.al. (1990).

The same warnings listed in paragraph 4.5.4 are also applicable here, meaning that the recommended guideline values should be dealt with on a case-by-case basis.

In terms of a BESD, $\hat{\omega}$ can be interpreted as follows: with all of the marginal totals equal to 100 and by using (5.62) , alter the top-left cell of the 2 x 2 – frequency table to:

$$100\sqrt{\hat{\omega}}\,/(1+\sqrt{\hat{\omega}}). \qquad (5.64)$$

Example 5.17:

(a) In Example 5.16 we found that $\hat{\omega} = 3,02$ and the 90% *CI* for ω was (1,333 ; 6,840). The population value of ω can thus vary from a small effect to a large effect. The top-left cell of a BESD in this case, has frequency:

$$100\sqrt{3,02}\,/(1+\sqrt{3,02}) = 63,5 \quad ,$$

giving the following 2 x 2 – table:

		Dislike snakes		
		No	Yes	Total
Fear of	Yes	63,5	36,5	100
Snakes	No	36,5	63,5	100
	Total	100	100	200

The counternull value of $\hat{\delta}_{OR}$ is $2\hat{\delta}_{OR}$ = $2 \cdot \frac{\ell n(3,02)}{1,81} = 1,22$. Thus, the Dus the counternull value of $\hat{\omega}$:

$$e^{1,81 \times 1,22} = 9,12\,.$$

Therefore, the odds ratio of 9,12 is as probable as OR = 1.

(b) For Example 5.10, if the samples are randomly drawn from the populations of individuals with heart disease and the population without heart disease, the 95% *CI* for the population *OR* is calculated as:

$$\ell n\left(\frac{78\times 61}{42\times 59}\right)\pm 1,96\sqrt{\frac{1}{78}+\frac{1}{59}+\frac{1}{42}+\frac{1}{61}}$$

$$=\ell n\left(1,92\right)\pm 1,96\times 0,07=0,652\pm 0,137=\left(0,515;0,789\right).$$

Finally, the 95% *CI* for ω is:

$$\omega_L = e^{0,515} = 1,674 \;\; ; \;\; \omega_U = e^{0,789} = 2,201.$$

In this situation the population value ω has a small to a medium effect.

□

5.5 Effect size of relationship between two nominal variables

A sensible measure of the degree in which cell frequencies in a two-way frequency table deviate from the expected frequencies, if one assumes there is no relationship, is (Cohen, 1969, 1977, 1988):

$$w=\sqrt{\frac{X^2}{N}} = \sqrt{\sum_{i-1}^{m}\frac{\left(f_i - v_i\right)^2}{Nv_i}} \qquad (5.65)$$

where f_i is the i^{th} cell's frequency;

v_i is the expected frequency of the i^{th} cell if no relationship is assumed;

and $m = IJ$, where I is the number of rows and J is the number columns in the frequency table.

Further, the chi-squared test statistic, X^2, is used to test for statistical significance relationship when a random sample is drawn.

The expected frequency of a cell is:

(row total $\times$ column total of the specific row and column of the given cell)/N, where N = sum of row totals = sum of column totals = total frequency.

Example 5.18:

In Example B of Chapter 3, the frequency table, where lecturers were removed:

		Male students	Female students	Total
Type	SJ	57(64,79)	79(71,21)	136
	SP	29(24,77)	23(27,23)	52
	NT	23(20,01)	19(21,99)	42
	NF	12(11,43)	12(12,57)	24
	Total	121	133	254

$$X^2 = \frac{(57-64,79)^2}{64,79} + \frac{(79-71,21)^2}{71,21} + \frac{(29-24,77)^2}{24,77}$$
$$+\frac{(23-27,23)^2}{27,23} + \frac{(23-20,01)^2}{20,01} + \frac{(19-21,99)^2}{21,99}$$
$$+\frac{(12-11,43)^2}{11,43} + \frac{(12-12,57)^2}{12,57}$$
$$=4,074$$

Thus $w=\sqrt{\frac{4,074}{254}} = 0,127$ □

The measure w can serve as an effect size index for measuring the relationship between two nominal variables (temperament type and gender in Example 5.18). It is clear that the more f_i differs from v_i, the larger the value $\left(f_i - v_i\right)^2/v_i$ becomes and, if for most of the cells there is a large difference, then X^2 should also be large. The size of X^2 is also influenced by N and so X^2/N is also a more reasonable measure. In the special case of 2 x 2-tables

$$\varphi^2 = X^2/N = w^2 \quad , \qquad (5.66)$$

which is also the reason why $\sqrt{X^2 / N}$ can be used as an effect size index for relationships.

Smithson (2000: 313) shows that while N influences the size of X^2, the number of cells also play a role in the sense that the more cells in the table, the larger the X^2 value becomes (the number of terms in the sum increases). To compensate for this, one can use *Cramer's* V (see also Cohen, 1969, 1977, 1988):

$$V = \sqrt{\frac{X^2}{N(k-1)}} \quad , \tag{5.67}$$

where $k = min(I, J)$.

In Example 5.18, the value $k = 2$ is used, because $I = 4$, $J = 2$, so that V has the same value as w.

Note:
For smaller tables, V and w are almost the same. In these cases w can be interpreted in much the same way as a correlation because it lies between 0 and 1. However, the same cannot be said of V for larger tables. For $k > 2$ the maximum value of V is smaller than 1, meaning that the size of the table has an influence on the value of V.

5.5.1 Estimation of w

When a random sample is drawn from a population, the effect size-index w can be estimated by the statistic $\hat{w}$ by making use of the sample frequencies.

For smaller samples $\hat{w}$ *underestimates* w and the bias for w^2 is approximately $\frac{(I-1)(J-1)}{n}$, where n is the sample size. (see Steyn, 2002).

Therefore, w should rather be estimated by:

$$\tilde{w} = \sqrt{\hat{w}^2 - \frac{(I-1)(J-1)}{n}} \quad , \qquad (5.68)$$

which is approximately unbiased for w.

Example 5.19 (Smithson, 2000):

Through the use of the Crosspatch-program of Smithson, the following frequencies are obtained from a random sample where the preferences of 10 to 40 year-old people divided into 3 age groups and into 4 types of shoes:

Shoe sizes

		1	2	3	4	Total
Age	10-19	86(44)	5(12,7)	38(54,6)	14(31,7)	143
	20-29	4(18,8)	14(5,4)	4(23,3)	39(13,5)	61
	30-39	14(41,2)	11(11,9)	87(51,2)	22(29,7)	134
	Total	104	30	129	75	338

$X^2 = 194,01 \ \left(p<0,0001\right)$

$\hat{w}=\sqrt{194,01/338} = 0,758$

$\hat{V} = w/\sqrt{2} = 0,536 \ \left(because\ k = 3\right)$.

There is a statistically significant relationship $\left(p < 0,000\right)$ in this example. The estimator's value is $\hat{w}$ = $0,758$, and it can be used to determine the effect of the relationship between shoe type and age in the population. This estimator is, for practical purposes, unbiased because the bias $\hat{w}^2$ is only approximately (2 x 3)/338=0,018, meaning that $\tilde{w} = \sqrt{0,758^2 - 0,018} = 0,746$ which is close to $0,758$.

5.5.2 Confidence interval for w

According to Johnson et al. (1995: 467) the chi-squared statistic X^2 has an approximate non-central chi-squared distribution with $(I-1)(J-1)$ degrees of freedom and non-centrality parameter $ncp = nw^2$. As in the case of δ in paragraph 4.1.2, it is possible, by making use of computer programs to first determine a $100(1-\alpha)\%$ *CI* for ncp and from there it is possible to obtain an approximate a *CI* for w. The SAS-program which can be used to calculate this is called VI_w and can be downloaded from this manual's webpage.

The 95% *CI* for w in Example 5.19 is: (0.640; 855) which means that the unknown population value, w, can vary between 0,64 and 0,86 with a probability of $0,95$.

5.5.3 Guideline values for w

Cohen (1969, 1977, 1988) provides guideline values for w in a table where w and corresponding Cramer's V values are provided for different values of k. Table 5.6 is an extract of this table and makes use of the relationship stated in (5.40).

Table 5.6

Values of w and corresponding V

w	k =2	3	4	5	6
0,1	0,1	0,071	0,058	0,05	0,045
0,3	0,3	0,212	0,173	0,150	0,134
0,5	0,5	0,354	0,289	0,250	0,224

Note that if $k=min(I,J)=2$, then $w=V$. When $I=J=2$, then we have that $w=\varphi$.

The guideline values chosen for w are thus based on the same guidelines used for φ:

- Small effect: $w = 0,1$
- Medium effect: $w = 0,3$
- Large effect: $w = 0,5$.

Cohen warns that for larger tables these guidelines might be unrealistic. Cramer's V is actually a modified index for larger tables and so Table 5.6 can then be used. For example, if $I = 6$ and $J = 10$, then $k = 6$ and V-values of $0,224$, $0,134$ and $0,045$ can be considered as large-, medium- and small effects.

In Example 5.18 we found that $w = 0,127$ and because $k = 2$ in Table 5.6, it is found to be a small effect. For a larger table with the same w-value it would be considered a medium effect if k was, for example, larger than 4. In Example 5.19 even the lower bound of the 95% *CI* gives us the right to classify it as a large effect (because for $k = 3$ it is a large effect $0,354$).